Probative Blindness and False Assurance about Safety

Andrew Rae and Rob Alexander

Abstract

Safety activities may provide assurance of safety even where such assurance is unwarranted. This phenomenon – which we will call "probative blindness" – is evident both in hindsight analysis of accidents and in the daily practice of safety work. The purpose of this paper is to describe the phenomenon of probative blindness. We achieve this by distinguishing probative blindness from other phenomena, identifying historical instances of probative blindness, and discussing characteristics and causes associated with these instances. The end product is an explanation of the features of probative blindness suitable for investigating the probative value of current safety activities, and ultimately for reducing the occurrence of probative blindness.

Keywords: Probative blindness, assurance, compliance, qualitative analysis, accident analysis

1 Introduction to Probative Blindness

Not all "safety activities" have a positive effect on safety. Some activities neither reduce the risk of harm associated with a system, nor provide more accurate understanding of that risk. These activities have no safety value. Worse, if these activities are believed to be effective, they result in false assurance – unjustified confidence that safety goals have been met.

Safety activities perform three main roles, where any given activity can fill more than one role. "Ensurance" is the direct improvement of the safety of systems or operations. An example of an ensurance-focused activity is changing a design by adding diverse means of performing a safety function. "Assessment" improves knowledge about safety. Quantitative Risk Assessment is the stereotypical assessment technique – it does not directly make a system safer, but is supposed to inform ensurance effort (Apostolakis 2004) and thus indirectly improve safety. "Assurance" is the demonstration of safety, often directed towards increasing the confidence of stakeholders not directly involved in ensurance and assessment. As with assessment, assurance does not directly make a system safer, but efforts to demonstrate safety may lead ultimately to safety improvement. For example, attempting to construct a formal proof that a design meets its safety requirements may expose ambiguity in the requirements or bugs in the design.

Assessment and assurance are closely linked. The difference between determining safety and demonstrating safety is subtle, with the terms "assessment" and "assurance" often used interchangeably. Regulatory approaches typically assume that an activity that demonstrates safety would equally reveal danger if such danger was present (Menon, Hawkins, and McDermid 2009). Where this is not the case the activity is an instance of "probative blindness" (Rae et al. 2014).

Any safety technique, applied in the wrong way or under the wrong circumstances, can exhibit probative blindness. Hence, probative blindness is a property of activities; it applies to particular executions of a technique by particular people at a particular time.

An activity is defined as exhibiting probative blindness if it provides stakeholders with subjective confidence in safety disproportionate to the knowledge it provides about real problems.

There are often multiple opportunities to identify and mitigate hazards, so isolated instances of probative blindness are not necessarily catastrophic. When an organisation is prone to probative blindness, however, its beliefs about safety may drift away from reality even as great effort is expended on safety activities. This is why probative blindness is of such concern – it involves substantial wasted effort, and it can actively hide problems. Probatively blind activities can engage skilled people in enthusiastically doing things that increase the risk of harm. This is a particularly galling misuse of good engineers with good intentions. A better understanding of the phenomenon is necessary if we are to build organisations that can select, apply and interpret safety activity to align beliefs about safety with safety reality.

In order to learn about probative blindness, we need a research approach matched to the current maturity of our understanding. Moving ahead too quickly – developing sophisticated theoretical models of loosely defined phenomena – can be unhelpful. Before a phenomenon such as probative blindness can be theorised, it must first be distinguished and explored (von Krogh, Rossi-Lamastra, and Haefliger 2012).

This paper presents a case study series designed to characterise historical instances of probative blindness. The case studies show how probative blindness can be distinguished from other phenomena, and provide an initial characterisation of the manifestations and causes of probative blindness.

The paper argues that probative blindness is a distinct and recognisable phenomenon, illustrates the features by which probative blindness can be recognised, and suggests how the causes of probative blindness can be investigated further.

2 Distinguishing Probative Blindness as a Phenomenon

2.1 Belief-shifts have a central role in accident theory

Organisational accident theory suggests that accident prevention hinges on early recognition that a dangerous situation is developing. In other words, there needs to be a shift from believing that the situation is "safe" to believing that the situation is "unsafe". The reasons behind this lack of belief-shift become a central theme of the accident narratives. For example, Weick (1993) described the deaths of thirteen fire jumpers in the Mann Gulch fire in terms of their understanding about how dangerous the fire was. Early impressions that it was a fire that could be extinguished by the next morning were reinforced by the actions of their team leaders. When (too late) they realised that they were in imminent danger, team co-ordination and trust collapsed. In the wake of the Hertfordshire Oil Storage Depot (Buncefield) explosion, the Health and Safety Executive criticised the operators for "not understanding the potential impact of a vapour cloud explosion" (Board 2006). The implication of this criticism was that the actual beliefs about danger differed from the "correct" beliefs about danger. The Royal Commission into the West Gate Bridge collapse, in discussing a particularly dangerous feature of the construction method, suggested "Neither contractor … appears to have appreciated this need for great care" (Barber 1971). A dangerous situation had developed without a corresponding change in the perception of risk.

Whilst individual accident reports will often make claims about things that "could have been known" or "should have been known", there will always be a mismatch between what appears obvious in hindsight and what was actually knowable with foresight (Fischhoff

2003). Incorrect beliefs that appear unreasonable to investigators probably were rational to those with no knowledge of what was to come. Attempts to provide a general theory of accidents, summarised in Table 1, try to reconstruct this rationality. In particular, they offer explanations for how and why beliefs do not shift to match the real safety of the system (which would have allowed operators or designers to prevent the accident).

Turner (1976) describes the pre-accident period as "disaster incubation". During disaster incubation the organisation does not shift its beliefs about safety despite mounting evidence of problems. Turner's explanation for this problem is a form of bounded rationality, where organisations are unable to pay attention to signals of danger. These signals are important and obvious in hindsight, but before the accident appear as insignificant – even as distractions from more salient concerns.

Subsequent researchers have upheld Turner's characterisation of the problem as a failure to shift beliefs, but have offered alternate explanations for how beliefs are formed, evolve, and are challenged within organisations.

Keyser and Woods (1990) describe the problem of "fixation errors". A fixation error involves a preliminary assessment of a situation that is rational given the information available at the time. This early assessment is not revised as new information becomes available, or even as the situation itself changes. Keyser and Woods provide the example of an operator ignoring alarms because they "know" that the alarms are inconsistent with the "actual" state of the system.

Vaughan (1997), explaining why the space shuttle program did not react to increasing evidence of danger, introduced the concept of "normalisation of deviance". Once a particular warning signal has become absorbed into routine operations, further occurrences of similar signals have no particular salience. Instead of suggesting that the state of affairs is unsafe, they are part of a pattern of information associated with a normal, presumed safe situation.

"Normal Accidents", written by Perrow (1999) in the wake of the Three Mile Island accident, suggests that the complexity of interaction between human and technical systems can render the current state of the combined human-technical system incomprehensible. People form flawed mental models of the system, and then interpret new information (which could have vital safety insights) to fit those models; they are unlikely to quickly update the models themselves in the midst of an emerging dangerous situation.

Kewell (2007) points to the role of reputation as a two-way "cloaking device", both concealing risk from outsiders and preventing insider awareness of danger. Strong existing beliefs, constructed through a process of public relations, institutionalisation and mystification, are resilient to new information, particularly when the source of that information is less socially powerful.

"High Reliability Organisations" (La Porte 1996) suggests that organisations are safest when they focus on "evidence that contradicts" and eschew hierarchical authority in factor of operational knowledge.

Table 1: Accident Theories Involving Belief-Shift

Theory	Author(s)	Primary Concerns
Disaster Incubation	Turner, Pidgeon	Bounded rationality, particularly for leadership attention and decision making
Fixation Errors	Keyser & Woods	Situation assessment by operational staff
Normalisation of Deviance	Vaughan	Differentiating warning signs from routine events at all levels of the organisation

Normal Accidents	Perrow	Situation assessment by operational staff
Reputation	Kewell	Interactions between staff with different levels of authority
High Reliability Organisations	La Porte, Weick, Rochlin, Roberts	Operational decision making

All of these theories make the counter-factual claim that accidents could be prevented if only organisations were better at updating their beliefs. Failure to do so is explained in terms of properties of the organisations – structures, attitudes, technologies, and reputations – but the theories do not directly examine the events in which beliefs fail to shift.

2.2 Probative blindness is one of several belief-shift phenomena

Probative blindness is not intended to be a new theory of organisational accidents. Instead, it is a clarification of one of the phenomena that must be explained by organisational accident theories. Organisation-level properties – structures, attitudes, technologies and reputations – give rise to specific events in the lead up to accidents. Accident theories, whilst they try to explain pre-accident events, seldom consider the individual events themselves in detail, preferring to focus on organisation-level theories.

The purpose of theories is to "predict and explain phenomena" (Bogen and Woodward 1988). Theories that explain the major phenomenon of organisational accidents do so by describing smaller phenomena as symptoms of organisation-level properties, and drawing causal relations between these properties and the accidents. Unfortunately, new theories often introduce different labels to describe or categorise previously described phenomena, rendering the phenomena less distinct. Von Krogh (2012) suggests that organisational research is prone to such "early over-theorising", attempting to explain and link phenomena before they have been adequately described. This leads to confusion as phenomena are described through different theoretical lenses, disguising and blurring the boundaries of the phenomena themselves.

We hope to avoid this fate for probative blindness by clearly delineating its boundaries. Other pre-accident phenomena – such as absent safety activities or beliefs that are held by one organisation but not by another – warrant similar treatment, but are not the topic of this paper. The definition of probative blindness is:
An activity that provides stakeholders with subjective confidence in safety disproportionate to the knowledge it provides about real problems.

There are three elements to this definition. Probative blindness requires:

1. a specific safety activity, conducted at a particular time;
2. an intent or belief that the activity provides new information about safety; and
3. no change in organisational belief about safety as a result of the activity.

2.2.1 A specific safety activity

Organisations use a range of techniques such as Hazard and Operability Studies (HAZOPS), Fault Tree Analysis (FTA) and Failure Modes and Effects Analysis (FMEA) to form and update their safety beliefs. Each activity – each particular instance when such a technique is applied - is capable of updating or not updating beliefs. Any activity that purports to provide information about safety, but does not change organisational beliefs about safety exhibits probative blindness.

Probative blindness is an outcome of activities rather than of techniques, organisations, or products. Of course, some techniques may be more likely to lead to probative blindness than

others, and some organisations may be more prone to probative blindness. If so, these are causes of probative blindness, not the phenomenon itself.

2.2.2 An intent or belief that the activity provides new information about safety

Many activities were never intended to change beliefs. We will use the broader category "non-probative" to encompass such activities. Most ensurance activities – those that seek to improve safety by directly modifying systems – have no probative intent (although in practice they tend to increase confidence of safety). Safety promotion activities (e.g. poster and training campaigns) fall into this category unless they also influence the beliefs of key decision makers. "Non-probative" is thus not, in itself, a pejorative term; it is only so when it is applied to an activity that is *intended* to be probative. This difference in intent makes "non-probative" a neutral observation about an activity, and "probative blindness" a judgement of failure.

Failure to update beliefs is not, by itself, probative blindness, because this failure could also come about through a lack of any safety activity. Activities that were never conducted cannot, by definition, be non-probative. Whilst there are accident narratives in which necessary safety activities were absent, this is a separate phenomenon from probative blindness.

Probative blindness may thus be confused with lack of willingness or effort to engage in safety activities. Whilst Turner (1976) explicitly prescribes safety analysis as a solution to cultural blindness, many major accidents are preceded by substantial safety management activity, including voluminous risk analysis. For example, prior to the Challenger explosion there was no reluctance within NASA to spend substantial effort on identifying and managing safety issues, but a strong reluctance to believe that there were real safety problems. (Vaughan 2004).

Probative blindness and absent safety activity are different phenomena that may appear in the same organisations and accident narratives. Within organisations that pay little attention to safety, some activities may be missed, and other activities that are performed may have low status or be under-resourced, leading to probative blindness.

2.2.3 No change in organisational belief

Organisational beliefs, for the purpose of defining probative blindness, are those that influence the behaviour of key decision makers. Examples of beliefs relevant to safety are:

- the overall level of safety of a physical system;
- the success with which a particular hazard has been mitigated;
- the risk associated with an operation; and
- the likelihood of a particular catastrophic event.

Correct beliefs (recognising that a safe system is safe, or appreciating the true amount of risk) are of course a good thing to have. Probative blindness is dangerous because it can lead to false beliefs. However, defining probative blindness in terms of correctness of belief is not useful, because the correctness or otherwise of beliefs about safety can usually only be judged in hindsight. Instead, we define it by the inability of an activity to cause any change in belief, whether in a more or less accurate direction. In practice, of course, we are indeed most concerned about beliefs that turn out to be dangerously wrong.

2.3 Probative blindness can be identified retrospectively

In order to identify probative blindness, it is necessary to specify where organisational beliefs exist, and how changes in these beliefs can be observed. Often it is not the beliefs of the

people directly performing the safety activity that matter, but the beliefs of those responsible for making decisions about designs and operations. These people receive and interpret the outputs of the safety activities, and it is their beliefs as decision makers that are relevant for identifying probative blindness.

After an accident, all evidence is interpreted with the benefit of hindsight. The two main types of evidence about beliefs are pre-accident actions (as recorded or related by witnesses) and post-accident testimony directly about beliefs. Testimony about beliefs is subject not only to hindsight bias, but also to rationalisation, as witnesses reconcile their pre-accident and post-accident beliefs. Both "I should have known" and "I could not have known" say more about the state of mind of the witness after the accident than before it.

Actions can provide more direct evidence of belief, but only if the researchers discard the possibility of wilful negligence. By definition, negligence involves action or inaction that is unreasonable in a particular situation. There is no way for hindsight of the actions alone to distinguish between reasonable behaviours based on incorrect beliefs, and unreasonable behaviours based on correct beliefs.

Characterising belief-shift through hindsight requires three things. Firstly, there must be a working assumption that people acted reasonably based on the beliefs that they held. Secondly, they must have exhibited behaviour that is consistent with one set of beliefs and inconsistent with alternative beliefs. Thirdly, this behaviour must have left an evidence trail accessible to researchers. These conditions will not be true in all cases, even where probative blindness existed. However, they are present often enough to allow useful research into the phenomenon.

3 Examples of Probative Blindness

3.1 Method for identifying probative blindness in accident narratives

Our project used a set of four case studies, where each case study was a single major accident. We selected these according to whether existing narratives of the accident include beliefs about safety as a significant factor in the period leading up to the disaster. The source material varied for each case study, but in each case included at least one official accident report, along with supporting documents such as interview transcripts and pre-accident safety analyses.

Within each case study we constructed a list of safety activities from the case study documents, and collated all mentions of each activity. We then used an alternate theoretical templates method, as described by Yin (2003), to determine whether each activity was an instance of probative blindness. This method matches case study observations to patterns of predictions associated with different theories or phenomena. Our intent in this study was to find safety activities that matched probative blindness but did not match other phenomena.

In order to follow this approach, it was necessary to make use of a number of auxiliary hypotheses. These are contestable statements that cannot be resolved within the study (given the methods we are using and the focus of our attention).

1. Accidents have proximate causal factors (PCFs) that existed within the system or organisation prior to the accident.
2. Prior to each accident, the organisation had collective beliefs about the existence and severity of the PCFs.
3. Hindsight after the accident provided improved knowledge about the existence and severity of the PCFs.

Further justification for these hypotheses will not be provided in this paper. They are necessary for any hindsight analysis of accidents (although seldom stated explicitly).

The definition of probative blindness from Section 2.2 requires:

- a specific safety activity, conducted at a particular time;
- an intent or belief that the activity provides new information about safety; and
- no change in organisational belief about safety as a result of the activity.

Additionally, for the purpose of this study we required that the activity was linked in the case study documents to a proximate causal factor for the accident.

The alternate templates we used were:

Activity not performed: Some activities described in accident reports are counterfactual – they are notable because they did not happen. In some cases they were required by standards or by simple good practice; in other cases, they were not explicitly required, but with the benefit of hindsight it's clear that they might have revealed important information. Counterfactual activities within the case studies were identified by direct reference to the fact that they did not take place, or by consistent reference to them as hypothetical activities (language such as "should", "would", and "if" used with reference to the activity taking place).

Neither blind nor safe: Some activities generate belief shift, but the beliefs do not translate into effective responses. In these instances, the activity may still be described as "ineffective" but the failure comes through ability or willingness to act, not through a lack of awareness. These activities were classified by finding language or actions subsequent to the activity that indicated increased awareness of the PCF.

Not intended to be probative: Some accident reports document activities that were not intended to be probative at the time they were performed. They may still in hindsight be labelled as "missed opportunities", but there is no direct evidence that there was intent or belief on the part of those who commissioned or performed the activity that it was capable of revealing a PCF. This template did not apply if the sole reason that an activity was expected to find nothing was that there was a belief that there was nothing to find, i.e. where the activity was expected to be probative and to legitimately reveal that there were indeed no problems.

Irrelevant: Accident reports occasionally discuss activities that are relevant for safety, but irrelevant for the particular accident that occurred. Such activities are used, for example, to indicate a generally good or generally poor standard of safety within the organisation.

By sequentially applying these templates to each activity, we were able to eliminate activities that had interpretations other than probative blindness. As shown in Table 2, the four alternate templates "Activity not performed", "Neither blind nor safe", "Not intended to be probative" and "Irrelevant" are each orthogonal to "Probative blindness" on exactly one criterion, which allowed us to match each safety activity to a unique template.

Table 2: Criteria Matching Alternate Templates

	Specific activity performed	Intended to be probative	No change in belief	Linked to accident
Activity not performed	NO	YES	YES	YES
Neither blind	YES	YES	NO	YES

nor safe				
Not intended to be probative	YES	NO	YES	YES
Irrelevant	YES	YES	YES	NO
Probative Blindness	YES	YES	YES	YES

The outcome was a set of activities for each case study that matched our definition of probative blindness.

3.2 Method for characterising probative blindness

We analysed the instances of probative blindness using a variant of Grounded Theory analysis. Grounded Theory is a "bottom up" approach to analysis of text, where the researchers begin by identifying and labelling ("coding") significant terms and ideas within the text. The codes are grouped into concepts, which in turn are assembled into larger patterns. There are several variants of Grounded Theory. Our work follows the approach of Corbin and Strauss (1990) who recommend a particular coding paradigm to aid with validity and replication. This paradigm looks for "causal conditions", "phenomena", "intervening conditions", "action strategies", and "consequences".

Our interest was in the real-world events prior to the accident, not the social construction of reality by the investigators. We interpreted statements in the transcripts and reports as a straightforward representation of real-world events, and accepted at face-value causal claims made in the accident reports.

For each activity that we had identified as probative blindness, we coded all references to that activity. The codes were initially terms selected directly from the text. These were grouped where different terms referred to the same concepts, and then structured into themes based on similarities between concepts.

Ultimately, the themes fell into two categories:

- the manifestations of probative blindness (i.e. the immediate mechanisms by which activity outputs fail to update beliefs)
- the conditions under which probative blindness occurs (i.e. things that cause probative blindness)

As noted above, Corbin and Strauss (1990) suggest that Grounded Theory should identify further categories relating to intervening conditions, action/interaction, and consequences. In the case of Probative Blindness, these categories would describe how the phenomenon could be avoided or managed. Such topics were weakly present in the accident reports, and were contained only in the recommendations. We decided not to focus on avoidance or management in this paper.

3.3 Case Studies

3.3.1 Boeing 787-8 Battery Incidents

On January 7, 2013, a Japan Air Lines Boeing 787-8 experienced a fire inside the Auxiliary Power Unit (APU) whilst parked at Logan International Airport, Massachusetts (National Transportation Safety Board 2014) . On January 16, 2013, an All Nippon Airways 787-8 experienced an in-flight main battery failure, resulting in an emergency landing and

evacuation (Japan Transport Safety Board 2014). Both events arose from a short circuit inside a lithium-ion battery cell, leading to thermal runaway cascading to adjacent cells.

Whilst there were no injuries in the first incident, and only minor injuries in the second incident, the incidents revealed a potentially catastrophic failure mechanism that was not reflected in the aircraft safety assessment.

Boeing had overall responsibility for integration and certification of the Electrical Power System (EPS). Thales designed the subsystem containing the main and APU batteries, which were supplied by GS Yuasa. Certification activities were overseen by the Federal Aviation Administration (FAA).

The following safety activities exhibited probative blindness:

Nail Penetration Test
GS Yuasa conducted a nail penetration test in November 2006. This test involved driving a steel nail through one of the battery cells in order to induce a short circuit and observe the effects. As a result of the single documented nail penetration test, GS Yuasa concluded that a short circuit in a single cell would not propagate to other cells or result in a fire (a conclusion eventually contradicted by the incidents described above). This belief influenced both the design of the battery enclosure and the quantitative risk assessment for the Auxiliary Power Unit.

Post-Assembly Inspection
GS Yuasa conducted physical and CT scan inspections of each manufactured battery. These inspections were intended to detect manufacturing defects such as foreign object debris (FOD) and wrinkles in the cell windings. Less than one percent of batteries were rejected during these inspections. After the accident, National Transport Safety Board (NTSB) investigators found that the resolution settings on the CT scan machines were such that FOD and wrinkles could not be observed. Disassembly of in-service batteries during the investigation revealed numerous defects and variances.

EPS Safety Assessment
Boeing's Safety Assessment for the Electrical Power System (EPS) included a quantitative risk assessment for the Auxiliary Power Unit, prepared by Thales. This included Fault Tree Analysis (FTA) and Failure Modes and Effects Analysis (FMEA). The safety assessment concluded that the likelihood of a battery cell venting with smoke was less than one in ten billion flight hours. Both incidents above occurred during the first 52000 flight hours, suggesting that the actual likelihood was considerably higher.

Boeing and FAA Oversight
Compliance with the battery safety requirements was demonstrated through formal analysis and testing. Boeing reviewed the analysis provided by Thales and GS Yuasa, but did not uncover the problems with the assumption about cell propagation or the related likelihood of battery venting with smoke.

Similarly, FAA representatives reviewed Boeing's complete safety analysis for the EPS without uncovering the problems.

3.3.2 Deepwater Horizon Blowout

The Deepwater Horizon mobile drilling rig was preparing an exploratory well at Macondo Prospect in the Gulf of Mexico. On 20 April 2010, as the Deepwater Horizon was preparing to cap and abandon the well, there was a "blowout" – an uncontained release of gas and drilling fluid. The blowout led to an explosion that sank the Deepwater Horizon. There were

11 fatalities, and one of the largest oil spills in history (Deepwater Horizon Study Group 2011).

Commentary on the accident has focussed on the failures of BP (owners of the Macondo Well) and Transocean (operators of the Deepwater Horizon) to manage the risks associated with a blowout.

The following activities exhibited probative blindness:

Cement Slurry Analysis
The final cementing operation used nitrogen foam cement. Samples of the cement were laboratory tested to ensure that it had the required properties – including viscosity, curing rate, and final strength – under well conditions. Two tests were conducted in February, and two in April. The earliest three tests revealed the slurry to be unstable under the test conditions, but this was ascribed to problems with the tests rather than problems with the cement. The final test was completed only after the cement was used in the well – i.e., too late to influence decision making on the Deepwater Horizon.

Negative Pressure Tests
To verify the integrity of the cement barrier, "negative pressure tests" were conducted. These tests involved reducing the pressure within the drill pipe to 0 psi. If the cement barrier was working, it should have prevented the high hydrocarbon pressure in the rock formation through which the drill shaft descended forcing oil or mud into the pipe. Either a flow of liquid or a change in pressure would have indicated that there was seepage of fluid into the well, and therefore a problem with the cement barrier.

The tests should have only passed if there was no flow or increase in pressure. During the first test, it was not possible to reduce the pressure fully in preparation for the test, and the pressure rose during the test. For the second and third tests the partial vacuum was successfully created, but the pressure increased during the test. Rather than accept that the tests had failed, the crew hypothesised a "bladder effect", i.e. that the drill pipe was being squeezed by pressure in the surrounding riser. If this were true, the increasing pressure would not have indicated a failed cement barrier.

A revised test was conducted using the "kill line", one of several smaller pipes used to circulate fluids in the well. This revised test was a valid process – in fact it was the procedure submitted in the most recent drilling permit – but it only made sense to perform it when the drill pipe and kill lines were at the same pressure. The drill pipe started and remained at high pressure throughout the test, indicating that the integrity of the cement barrier was compromised. Nevertheless, the crew interpreted the constant low pressure in the kill line as a successful test, and evidence of the integrity of the cement barrier.

Blowout Preventer (BOP) Safety Analysis
The blowout preventer was subject to intensive safety analysis. The control system alone was the subject of a 472 page quantitative risk assessment (EQE International 2000). A detailed analysis of the blowout released in June 2014 determined that the true cause of BOP failure (the pipe was hard to shear due to buckling) was not only not considered in the risk assessment, but was likely to be missing from most analyses of similar systems (U.S. Chemical Safety and Hazard Investigation Board 2014). Transocean's Major Hazard Risk Assessment for Deepwater Horizon showed a clear belief in the efficacy of the blowout preventer (U.S. Chemical Safety and Hazard Investigation Board 2014).

3.3.3 Space Shuttle Columbia Explosion

The Columbia accident occurred on flight STS-107, February 1 2003. During ascent, a large piece of insulation foam, the Left Bipod Foam Ramp, was shed from the external fuel tank

and struck the leading edge of the orbiter wing. This damaged the Thermal Protection System, the heat-resistant tiles intended to protect the orbiter during re-entry to the atmosphere. Despite some concern during the mission, it was judged safe to attempt a landing, during which the orbiter disintegrated.

The original design requirement for the shuttle was that debris should not shed from the external tank at all. During the very first Columbia flight, which was the first flight of any shuttle, large amounts of debris were observed. On flights STS-7, STS-32R, STS-50 and STS-112 the Left Bipod Foam Ramp was observed as debris. Not all of these flights involved Columbia, but all involved the same design of external tank.

Throughout the shuttle program numerous safety assessments were conducted which served to alleviate, rather than reinforce, concern about the foam debris in general, and the Left Bipod Foam Ramp in particular.

The Columbia Accident Investigation Board Report (2003) contains some language that supports the "neither blind nor safe" interpretation of these assessments, suggesting an irresponsible acceptance of known risk rather than a failure to update beliefs. However, there is no direct evidence in the report or the accompanying documents to match this template. The outcome from each activity listed here was renewed confidence that the risk from foam debris was low, and that the shuttle was fundamentally safe.

The following activities exhibited probative blindness:

STS-8 Pre-Flight Review
During flight STS-7 the Left Bipod Foam Ramp struck the Thermal Protection System. This event was designated after the mission as an in-flight anomaly, a classification that required proof that there was no threat to flight safety. The STS-8 pre-flight review that "resolved" the anomaly was an instance of probative blindness. This review discussed repair of the damage, but did not refer to future risk.

Analysis between STS-35 Post Flight and STS-36 Pre Flight Review
After flight STS-35 considerable damage was observed to the Thermal Protection System. This was designated an in-flight anomaly. The anomaly was closed by the STS-36 pre-flight review, which stated that there was no increase in Orbiter Thermal Protection System damage compared to previous flights and that it was 'not a safety-of-flight concern.'

STS-50 and Integration Hazard Report 37
During flight STS-50 the foam ramp detached and caused what was, at that time, the largest ever area of tile damage. The incident was recorded as "Integration Hazard Report 37", which was closed out as an "accepted risk". The close out activity is not described. It is this activity which exhibited probative blindness.

Pre-Flight Review STS-113
The briefing slides for STS-113 (which occurred before STS-107) discussed the foam loss on STS-112. These slides concluded "The ET [external tank] is safe to fly with no new concerns (and no added risk)". An accompanying report indicated a 99 percent probability of no foam being shed from the same area. This calculation was based on a selective sample of flights (i.e. it excluded several of the known Left Bipod Foam Ramp losses), inappropriately averaged between the left and right ramps, and assumed incorrectly that "no record" was positive evidence of "no loss". There were many occasions when the ramp condition was not observed or recorded. A review as part of the subsequent accident investigation estimated a 10 percent probability of Left Bipod Foam Ramp loss.

Debris Assessment Team during STS-107

During the second day of the STS-107 mission, review of the launch photos revealed a debris strike on the left wing of the orbiter. A "Debris Assessment Team" was formed to analyse the damage. This team used a modelling tool known as "Crater" to estimate the amount of damage from the size of the observed strike. The team requested that imagery be sought from outside sources to reduce the uncertainty of this estimate, but this request was not passed on. The Crater tool indicated that a dangerous amount of damage had occurred. However, the tool was believed to be conservative, and so the analysis team applied engineering judgment to conclude that the damage was acceptable.

The Debris Assessment Team was not highly confident of their own findings, and there is some evidence that the uncertainty caused considerable safety concern for individuals and the team as a whole. Their activities, despite the problems with the tool, were locally probative. However, as the results were transmitted to and interpreted by the mission management team, the uncertainty decreased, and the analysis was used to confirm belief in the insignificance of the damage. The activity caused no change of belief amongst the operational decision makers.

3.3.4 Australian Government Home Insulation Program Deaths

The Home Insulation Program was part of a stimulus package announced by the Australian government on 3 February 2009 (Hanger 2014). The program involved government-subsidised installation of insulation into private homes. As an intended effect of the program, many new contractors and employees entered the home insulation industry.

Between October 2009 and February 2010 four fatalities and two serious injuries occurred on separate projects funded by the program. All of the fatalities involved a lack of training and supervision for the installers, who were young men new to the insulation industry. Two of the fatalities were electrocutions caused by the use of metal staples in combination with reflective foil laminate (RFL).

The following safety activities exhibited probative blindness:

Industry Consultation Meeting

A formal consultation meeting was held on 18 February 2009, between members of the program office and representatives of the installation industry. No representatives of electrical trades were included in the meeting; however, training and competency were discussed, with direct links drawn between training and worker safety. The specific risk of electrocution related to RFL was also raised. The actions arising from this meeting did not reflect any increased awareness of the risks on the part of the program office staff. In particular, industry representatives presented information regarding deaths involving RFL installation in New Zealand. The fact that project staff did not seek more information about these deaths after the meeting is a strong indication that their beliefs about the risks had not been changed by the information.

Risk Register Preparation

On 3 March 2009 a risk identification workshop was held by the program office. On 13 March, Minter Ellison Consultants were appointed as external risk consultants, and a further risk identification workshop was held on 23 March. A draft risk management plan was produced on 30 March, and discussed at meetings on 31 March and 2 April. The output of this process was a risk register finalised on 9 April. During this same period, the Project Plan for the Home Insulation Program was being developed. The various versions of the Project Plan provide evidence of the evolution of understanding of risk within the team. Rather than improving understanding of safety risks, the risk workshops appear to have de-emphasised concerns about training, competence and workplace safety, ultimately not including these topics in the project risk register. The risk assessment effectively destroyed previous awareness of a safety issue. We might call this a case of retrograde probative amnesia.

Audits and Inspections
The audit and compliance scheme occurred in two phases. During the first phase, from the start of the program until October 2009, less than one percent of the installations were inspected. The focus was on detecting fraud rather than checking quality or safety of the installations, so desktop audits - examination of the paperwork - were prioritized. From October 2009 a new auditor was appointed, with more technical inspections, but still with a focus on detecting fraud and checking that homes were eligible for the program.

October 2009 Briefing Notes
After the first fatality there was a period of intense activity, including briefings for the responsible Minister delivered on 19 October, 21 October, and 22 October. These briefings specifically covered the RFL-related electrocution hazard, but recommended against removing RFL from the program. As with the risk register, this safety activity took as input a number of sources of information that clearly understood the risk, and produced outputs that understated and obscured the risk. Based in part on these briefings, RFL was not removed from the program for several months.

4 Manifestations of Probative Blindness

The instances of probative blindness within the case studies exhibit a small number of observable mechanisms. These mechanisms are observable only in hindsight – this is not a spotter's guide that can be used to observe probative blindness as it happens – but we hypothesise that the different mechanisms may have distinctive patterns of causation that can ultimately be recognised and remedied before they cause a problem. Each mechanism operates at a different stage during the activity. The categories and subcategories of mechanism are shown in **Error! Reference source not found.** and described in sections 4.1 to 4.4.

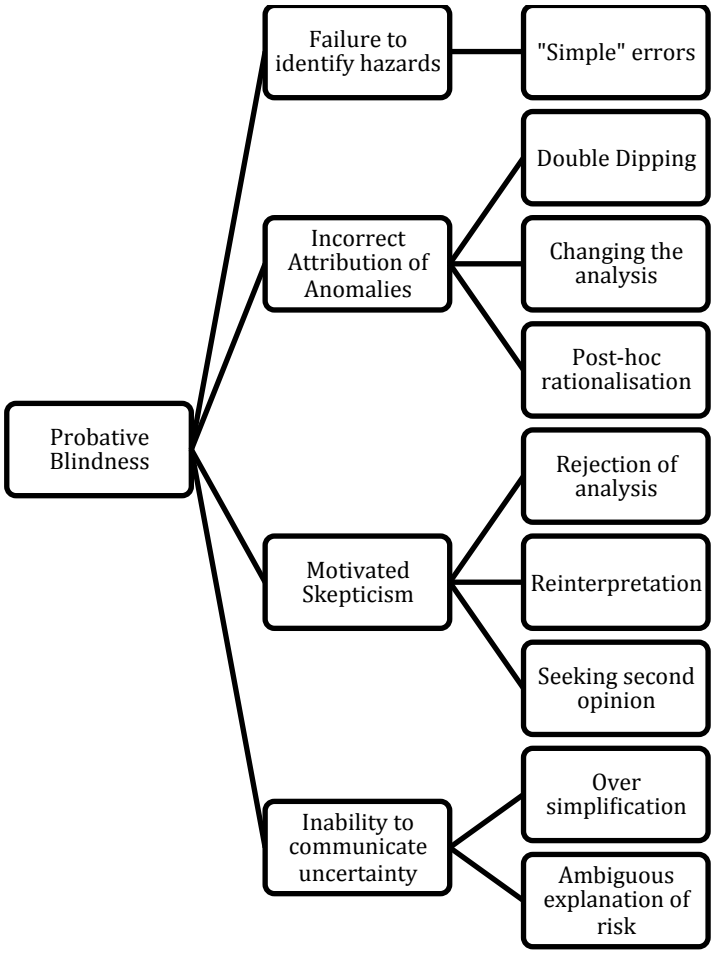

Figure 1: Manifestations of Probative Blindness

4.1 Failure to identify hazards or to correctly assess their significance

The first manifestation of probative blindness is "simple error" in the conduct of safety analysis. The analysis has the shape and form of a correct analysis, but produces outputs that do not recognise or pay sufficient attention to the problem or condition which ultimately leads to an accident. This may be failure to identify a hazard – (Barton and Rae (2012) suggest that approximately 20% of accidents involve failure to identify hazards in circumstances where it would be reasonable to expect the safety activity to do so) - or mistaken belief that a hazard has been adequately addressed.

From the case studies, this category includes:

- The 787-8 nail penetration test, EPS Safety Assessment and review
- The Columbia activities prior to flight STS-107
- Some but not all of the Home Insulation Program industry consultations; the risk register preparation; and the early audits and inspections

"Simple errors" have complex causes. Whilst the mechanisms of probative blindness are contained within a short time frame during the conduct of the activities, the causes and conditions stretch well beyond. In each of these cases it is likely that different people put into the same situation would have exhibited the same errors.

Discussion of the precise types of error that may be made during a safety analysis is beyond the scope of this work. For a detailed discussion see Rae, Alexander and McDermid (2014).

4.2 Incorrect Attribution of Anomalies

Sometimes the analysis team successfully discovers problems, but these are interpreted as mistakes in the safety activity. This manifestation has three subcategories.

1. "Double dipping" – repeating unfavourable tests until a favourable result is achieved. This was evident in the negative pressure tests and cement foam analysis for Deepwater Horizon. In each case, test failure was attributed to the test conditions rather than to the cement, and the test was repeated under changed conditions.

2. Changing the analysis or its interpretation until it produces a favourable answer. This requires a known "acceptable" answer, such as risk acceptance criteria. This was evident in the shuttle Debris Assessment Team's use of the Crater tool.

   "Again, the tool was used for something other than that for which it was designed; again, it predicted possible penetration; and again, the Debris Assessment Team used engineering arguments and their experience to discount the results." (Columbia Accident Investigation Board 2003, 168)

   In fact, the team did ignore the results – they decided that the tool was overly conservative, and produced a final damage estimate smaller than the one provided by the tool.

3. The third subcategory is post-hoc rationalisation – moving the goalposts for the activity to judge the results to be acceptable. In these cases, if there had been objective criteria fixed in advance, these criteria would have demanded an unfavourable result.

   The NTSB report into the Boeing 787-8 fires bemoaned the lack of objective criteria for battery inspections and thermal design features (National Transportation Safety Board 2014, 73). The final negative pressure test on Deepwater Horizon also involved post-hoc rationalisation, since strict application of the test criteria would not have judged the test to be passed.

4.3 Motivated Skepticism in the Interpretation of Bad News

Studies of motivated skepticism show that "information consistent with a preferred conclusion is examined less critically than information inconsistent with a preferred conclusion" (Ditto and Lopez 1992). Whilst no-one has directly investigated motivated skepticism in interpreting safety information, accidents such as the Challenger explosion (Vaughan 1997) and the West Gate Bridge collapse (Barber 1971) involved the rejection of unfavourable risk assessments by applying strict standards which were not equally applied to favourable risk assessments.

The analysis team may report unfavourable results, only to have the quality or suitability of the activity called into question. Critical review of safety activities is a good thing, but not if it is applied selectively to challenge uncomfortable results, but not confirmatory results.

This category includes outright rejection of the analysis; re-interpretation of the results in a way not intended by those who performed the analysis; as well as seeking a second opinion, and giving that more weight without any technical reason to prefer it to the initial judgement;

The following occurred during the space shuttle program.

Rejection:

"Program managers required engineers to prove that the debris strike created a safety-of-flight issue: that is, engineers had to produce evidence that the system was unsafe rather than prove that it was safe" (Columbia Accident Investigation Board 2003, 172)

Reinterpretation:
"In all official engineering analyses and launch recommendations prior to the accidents, evidence that the design was not performing as expected was reinterpreted as acceptable and non-deviant, which diminished perceptions of risk throughout the agency." (Columbia Accident Investigation Board 2003, 196)

The evidence from the pre-flight safety review process strongly suggests that technical judgements were being "overwritten" by second opinions. This was certainly the case in the earlier Challenger accident (Dombrowski 1991) and is a credible explanation for the otherwise inexplicably "sleight-of-hand" (Columbia Accident Investigation Board 2003, 126) calculations presented at the STS-113 Pre-Flight review.

4.4 Inability to Clearly Communicate Uncertainty

The fourth type of probative blindness concerns activities which are probative in themselves, but where the probative value is "lost in translation". This may be the case where uncertainty is felt and expressed by the analysis team, but is lost through over-simplification or ambiguous explanation.

The Home Insulation Program included numerous examples where risks were described in over-simplified terms, leading to misinterpretation of their intended scope. Worries about fraud, originally intended to include problems with competency, installation quality and certification, were transformed into desktop audits concerned only with financial fraud. Substantive discussions about safety, including the New Zealand experiences, were summarised in short line items about training (Hanger 2014, 93).

The Columbia Accident Investigation Board "found that a large number of hazard reports contained subjective and qualitative judgments, such as "believed" and "based on experience from previous flights this hazard is an 'Accepted Risk'". This led to an inability to share information about risk clearly throughout the complex organisation hierarchy (Columbia Accident Investigation Board 2003, 189).

The board also criticised safety reports that did not include "a quantifiable range of uncertainty and risk analysis" (Columbia Accident Investigation Board 2003, 168) – i.e. providing a range of probabilities rather than a single number. Where there were attempts to communicate uncertainty "Engineers … indicated a belief that management focused on the answer – that analysis proved there was no safety-of-flight issue – rather than concerns about the large uncertainties that may have undermined the analysis that provided that answer." (Columbia Accident Investigation Board 2003, 160).

5 Conditions under which probative blindness occurs

In this section we offer our own thoughts, supported by themes from the case studies, for why probative blindness occurs. Our claims about causality, particularly since they arise from hindsight and re-interpretation, are presented as cautious hypotheses rather than proven conclusions.

Any isolated instance of probative blindness, when viewed with the benefit of hindsight, gives the appearance of incompetence or misconduct. Instances are typically described as

"missed opportunities" – as if an activity that could and should have been probative was performed inappropriately.

By analysing the common conditions of a range of instances of probative blindness, however, our study suggests that activities that suffer probative blindness have a broader social and organisational role that actively subsumes and thwarts any probative potential. Whilst individual attitude and competence has a role to play, it is as much a symptom as it is a cause of probative blindness – individuals may have performed badly because the organisation set them up to do so.

Most safety activities that are not direct ensurance have a dual purpose assessment / assurance role. Whilst the activities have names typically associated with investigation, such as "assessment", "analysis", and "test", they also serve to communicate, demonstrate and reinforce existing beliefs. Safety activities are risk enablers as well as risk controls.

This dual purpose is normal and necessary. Communication and demonstration of safety is enshrined in safety regulation, and is demanded by customers and senior management. However, there is always a risk in this arrangement; a risk that participants lose sight of the fact that these activities actually could detect danger. The activities become entirely about assurance, and not at all about assessment. Even plans and schedules are determined on the assumption that the safety activities will never find problems.

Each of the case studies represents a large, public, expensive endeavour. The Space Shuttle was the flagship of the US space program. The Home Insulation Program was intended to be both a major public works initiative and a response to the global financial crisis. The 787 was an attempt to recapture the initiative in the Airbus/Boeing struggle for domination of the international commercial aircraft market. Deepwater drilling in the Gulf of Mexico was an ambitious use of novel technology and methods to restore US oil independence.

Projects such as these have corporate and political momentum. Whilst at face value regulations and safety analysis exist as checkpoints and controls, in practice they are a small part of a large and complex project execution process. They are gears being spun by the project machine, rather than levers with the power to direct and halt forward progress.

We make no claim that there is a causal relationship between "large public project" and probative blindness. Significant disasters were selected as our case studies because of the depth of investigation and the accessibility of the evidence - we did not compare probative blindness on large and small projects. However, these projects exaggerate and highlight the pressures present in all projects.

When a safety activity impedes a project it is an embarrassment and a failure. It is easier for a safety analyst – particularly if inexperienced and lacking the confidence that comes with deep competence – to believe that there is a problem with the safety activity than to accept that there is a problem with the project. It is easier still never to find problems at all.

The seven main conditions that we suggest lead to probative blindness are summarised in Figure 2 and elaborated in Sections 5.1 to 5.7

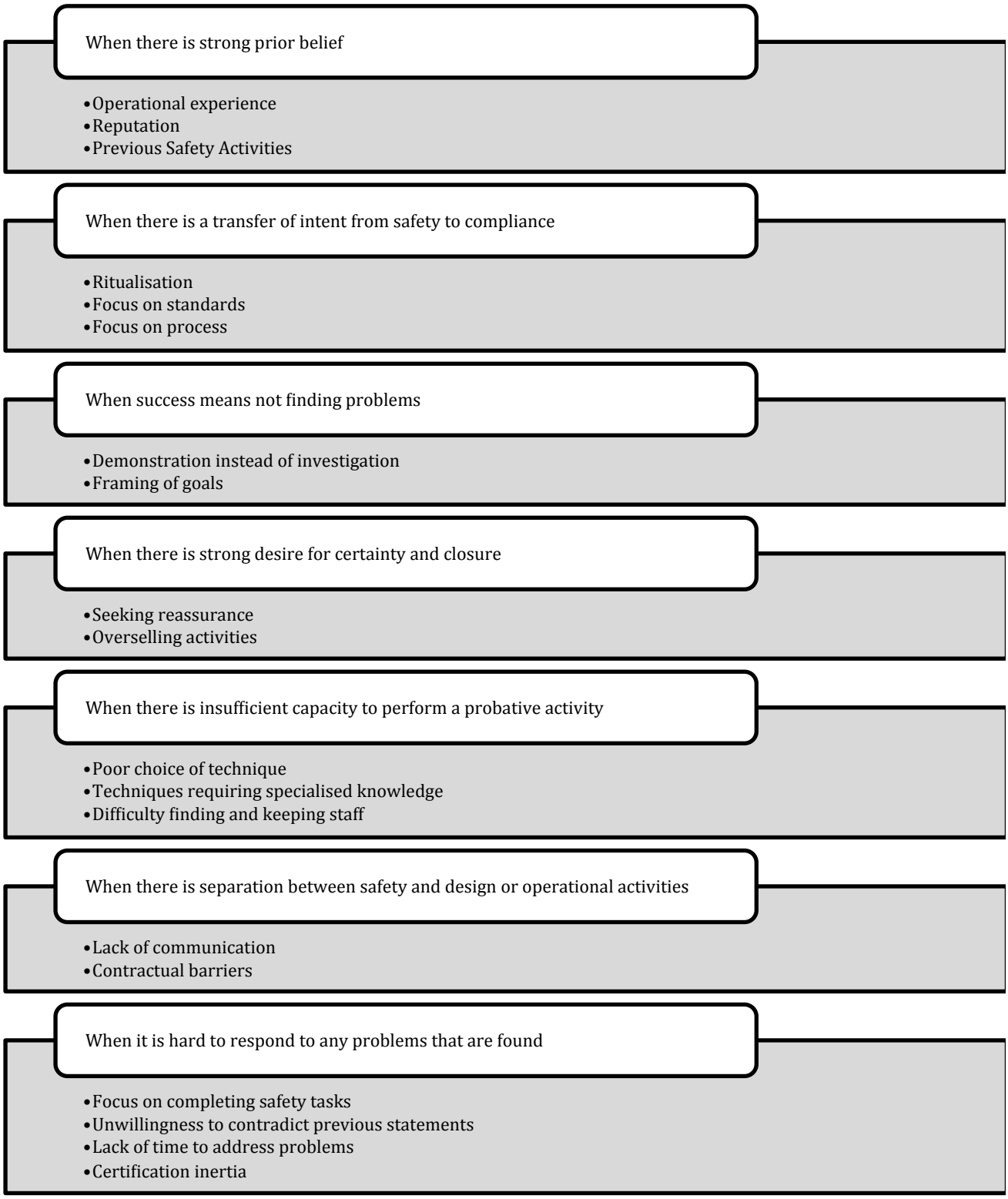

Figure 2: Conditions that give rise to probative blindness

5.1 Blindness can stem from strong prior belief in safety

"Belief" refers to the degree of certainty that existed prior to the safety activity. How much did the organisation, audience or analysis team expect the safety activity to find? Belief influences both the conduct and the interpretation of safety activities. When the team

conducting an activity does not expect to find problems, they are likely to conduct activities with mindsets and methods aimed at confirming information rather than testing it. Even where anomalies are found, there are always two possible explanations: either there is a problem with the thing being assessed, or there is a problem with the assessment. Which explanation is favoured is strongly influenced by which one is seen as more likely based on current belief, and on how acceptable each explanation is.

Each of the four case studies involved safety analysis conducted with a prior expectation of safety. This expectation was formed from operational experience, the reputation of the team responsible for the system, and even from previous safety activities.

At the time of the 787-8 safety assessment, GS Yuasa had over 14,000 lithium cells in service, all of similar design. None of these had experienced thermal runaway (National Transportation Safety Board 2014, 71). Armed with this "knowledge" that thermal runaway was highly unlikely, only a small number of physical tests were conducted, with only one nail penetration test properly documented (National Transportation Safety Board 2014, 70). An inverted pyramid was built upon these inadequate tests. The Thales FTA and FMEA used (and thereby endorsed) the GS Yuasa figures. The Boeing analysis used (and thereby endorsed) the Thales analysis. The FAA inspectors were not reviewing the safety of the batteries in a vacuum, but in the context of prior findings of safety by GS Yuasa, Thales, and Boeing.

On Deepwater Horizon, there was a ceremony aboard the rig on the day of the accident to celebrate seven years without a lost-time safety incident (Deepwater Horizon Study Group 2011, 38). The foam cement analyses and negative pressure tests themselves came after numerous prior tests and analyses. For example, the BP Accident Investigation Report makes frequent reference to "OptiCem", a software modelling tool used frequently throughout the placement of the cement. Whilst the limitations of this tool were recognised in hindsight, it is clear that significant reliance was placed on the tool during and immediately after cement placement. The prior findings of safety set the context for the negative pressure tests. The team conducting the tests believed that they were performing a routine check, and they expected a favourable result (BP 2010).

"Finally, due to poor communication, it does not appear that the men performing and interpreting the test had a full appreciation of the context in which they were performing it. Such an appreciation might have increased their willingness to believe the well was flowing. Context aside, however, individuals conducting and interpreting the negative-pressure test should always do so with an expectation that the well might lack integrity." (National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling 2011, 119)

When foam debris was first observed on the shuttle missions, it was classified as an "in-flight anomaly" – a serious event requiring extensive investigation. By STS-107, there had been 111 successful missions. Not only that, there had been multiple safety analyses directed specifically at the foam debris and tile damage issues. This created a considerable weight of expectation that any further analysis should reach a matching conclusion.

For example, the STS-107 Flight Readiness Review was based in part on the conclusions of the STS-113 Flight Readiness Review. This in turn used "Integration Hazard Report 37", from STS-50 (Columbia Accident Investigation Board 2003, 126) . Rather than an in-depth analysis, this "report" appears to be a line item in a hazard log concluding that the risk of foam debris strike was acceptable. The report was thus "ineffective as a decision aid" (Columbia Accident Investigation Board 2003, 187).

The Columbia Accident Investigation Board, discussing the in-flight analysis during the fatal mission, stated:

"It was at this point, before any analysis had started, that Shuttle Program managers officially shared their belief that the strike posed no safety issues, and that there was no need for a review to be conducted over the weekend" (Columbia Accident Investigation Board 2003).

Whilst in fact a preliminary damage assessment was conducted the following day, it was conducted with a strong prior expectation that any damage was tolerable.

The role of belief is less straightforward in the Home Insulation case study. There is no indication that there was specific belief in the safety of insulation, or in Reflective Foil Laminate (RFL) in particular. RFL had been banned in the New Zealand program due to electrocution risk, and there were numerous non-fatal electrical incidents involving insulation installation in the Australian state of Queensland (Hanger 2014, 76 & 195). However, workplace health and safety in Australia is primarily regulated at the State (regional) level, and there was a strong belief that existing State-based safety regulation was sufficient to address any safety risks. This belief framed all of the Federal Government discussions and assessments of safety.

"… 'assuming' that the States and Territories would monitor and enforce OH&S obligations, without taking any appropriate steps to ensure that such assumption was reasonable." (Hanger 2014, 223)

This analysis, which dismissed the risk based primarily on issues of scope and responsibility, was portrayed as a "Comprehensive Risk Assessment", and used to reassure the Minister that the safety risk was being appropriately managed. Throughout the subsequent Royal Commission, representatives of Minter Ellison maintained that they were not experts in the risks themselves, and were merely facilitating a process where risks were identified by the Department. Department representatives maintained that they were not experts in the risks, and were engaging Minter Ellison as external experts. (Hanger 2014, 141)

"DEWHA appointed a full-time Risk Manager for the HIP. That person's role was to coordinate and report on the risk management process and to provide feedback. The role was not to manage each individual risk, but rather to manage the risk framework. The function was primarily administrative rather than substantive risk management." (Hanger 2014, 148)

5.2 Blindness can stem from transfer of intent from safety to compliance

"Intent" refers to the reasons why a safety activity is carried out. What did the organisation, audience or analysis team want from the activity? In particular, were they seeking to learn about safety, or to reassure themselves or others that safety already existed?

If certain practices are known to improve safety, then it makes sense to require those practices to be performed, and to explain accidents in terms of an absence of those practices. Hence, there is a close link between "safety" – achieving an acceptable level of risk – and "compliance" – performing an accepted set of safety practices. Safety and compliance are closely linked in standards, regulation, and the approach of many organisations to safety (Besnard and Hollnagel 2012).

However, it is much easier to demonstrate that safety activities have been conducted than to demonstrate that they have improved safety. This can result in a transfer of concern from safety towards compliance for its own sake. Often, safety activities have an explicit goal of compliance or regulatory approval rather than the discovery or characterisation of risk.

Wastell (1996) describes how methods can become "an irrational ritual, the enactment of which provides designers with a feeling of security … at the expense of real engagement with the task at hand".

The transference of concern can be seen in management interest in the completion of risk activities rather than the details of the risks themselves. It can also be seen in the selection of specific activities based on regulations and standard practices rather than on a match between the activity and the risk under investigation.

The word "compliance" occurs 58 times in the NTSB report on the January 2013 Boeing 787-8 fire (compared to 18 mentions for the work "risk"). Almost all mentions refer to activities that were found to be compliant before the accident, but with the benefit of hindsight were inadequate.

"Critical assumptions and conclusions made in GS Yuasa's and Thales' safety analyses and used in Boeing's EPS safety assessment were not fully delineated and justified with appropriate data and engineering rationale. However, multiple independent reviews of the EPS safety assessment by Boeing authorized representatives and FAA certification engineers did not reveal these deficiencies." (National Transportation Safety Board 2014, 72)

Concern with risk management processes (rather than the risks themselves) was also evident in the Royal Commission into the Home Insulation Program.

"MR WILSON: Did you think—didn't you think you should find out what the residual risk was and how high it was?

[MINISTER]: No, I didn't because there's a sentence that follows that which refers to what was being proposed to particularly manage those risks.

MR WILSON: But in order to make sure that what's said in that sentence came to pass didn't you need to know what the residual risks were?

[MINISTER]: Well, that would be a matter for my Department and advisors to highlight for me if they determined it to be necessary. It would also be a matter for me to enquire for if I determined it to be necessary, but on the basis of this brief and the way in which this brief was expressed, no." (Hanger 2014, 122)

The Deepwater Horizon negative pressure tests were the sole activity conducted to test the integrity of the cement barrier. Additional techniques were available, in particular a "cement bond log" (CBL). A CBL team was standing-by on the drilling platform, but was flown back to shore after a decision was made to proceed with negative pressure tests only. This decision was made based on a pre-determined decision tree outlining the circumstances under which each activity would be performed (Deepwater Horizon Study Group 2011). The decision tree was in theory risk-based, but was too generic to consider all of the local conditions. Concern was transferred from selecting techniques based on indications of risk, to selecting them based on compliance with the decision-making tool.

The Columbia Accident Investigation Board does not use the specific term compliance, but does refer to managers being concerned more about the conduct of safety activities than with the content and outputs of those activities, and of an institutional faith in the efficacy of "bureaucratic accountability".

"Prior to Challenger, the can-do culture was a result not just of years of apparently successful launches, but of the cultural belief that the Shuttle Program's many structures,

rigorous procedures, and detailed system of rules were responsible for those successes." (Columbia Accident Investigation Board 2003, 199).

The board noted that any cultural improvement after the 1986 Challenger Accident had been subverted by renewed faith in the changed procedures and institutions.

5.3 Blindness can arise when "success" for the analysis activity means not finding problems

When a safety activity has a dual assessment / assurance role, subtle differences in context can switch it from an activity that primarily aims to investigate safety to an activity that primarily aims to demonstrate safety.

Both activities may produce evidence of safety, and both may discover safety problems, but the difference in goals leads to different choices and interpretations. The Haddon-Cave report in to the Nimrod XV230 accident recommended that "safety cases" be renamed "risk cases". Haddon-Cave's suggestion ignored the complication and confusion inherent in re-badging common practice, but made an important point about the relevance of language and intent for the probative power of safety activity. This was the same point made in the International Atomic Energy Agency response to Chernobyl, where they pointed out that carrying out an activity is not the same thing as carrying it out with the intent and willingness to find problems (International Nuclear Safety Advisory Group 1991).

One mechanism for framing "success" of safety activities is through risk acceptance criteria. The 787-8 safety analysis was performed against strict quantitative targets. The battery fires incident report states that "probabilistic methods demonstrate compliance in the certification process" (National Transportation Safety Board 2014). Demonstration, rather than discovery, was the purpose of the analysis.

NASA similarly maintained a safety program with a focus on demonstration. The accident board, referring to the pre-flight briefing slides for STS-113 stated

"This calculation was a sleight-of-hand effort to make the probability of bipod foam loss appear low rather than a serious grappling with the probability of bipod ramp foam separating." (Columbia Accident Investigation Board 2003, 126).

Even during Columbia's fatal flight, the Mission Management Team was concerned with the impact that the foam strike would have on the safety rationale for subsequent flights. In other words, foam strike was seen as a significant threat to the demonstration of safety, but not a threat to actual safety. A "successful" debris assessment team analysis would be one which addressed the threat to safety demonstration (Columbia Accident Investigation Board 2003, 147–48).

5.4 Blindness can stem from a strong desire for certainty and closure

Media commentary on all four of the accidents has suggested a lack of interest or concern for safety. However, none of the subsequent investigations produced evidence that this was the case. On the contrary, in all four cases senior managers expressed concern about safety, and sought reassurance. The organisational response was to provide the assurance they sought.

Whilst lithium batteries are used in many applications, their use in the main and auxiliary power systems of the 787-8 was novel. The Federal Aviation Administration produced nine "special conditions" for the use of the batteries. In essence, these conditions required that any

type of battery failure must be shown to be "not harmful" or "extremely remote". These requirements are typical of civil aviation regulation in that they placed considerably more emphasis on the base requirement than on the means of demonstration. The FAA had strong concerns about the safety of the batteries, but was willing to have those concerns assuaged by relatively weak evidence. (National Transportation Safety Board 2014).

The project office responsible for the Home Insulation Program was keen to portray a picture of good management and inter-departmental co-operation.

"Perhaps this was an effort by [the assistant secretary] and others to not tell the Minister of the tension between the OCG and DEWHA [the government departments with co-responsibility for the program]*, and to convey the impression that the Department very much had matters in hand. Of course, that is not what was in fact occurring."* (Hanger 2014, 138)

Briefing notes on the project referred to a "comprehensive risk assessment" to "identify and manage the full range of risks".

"The Minister was not told that the 'comprehensive risk assessment' was merely a facilitative process where the risks were identified by the Department and not by an engaged expert" (Hanger 2014, 141).

The internal NASA briefings during the Columbia flight also tried to provide clear answers and reassurances.

"Engineers who attended this briefing indicated a belief that management focused on the answer – that analysis proved there was no safety-of-flight issue – rather than concerns about the large uncertainties that may have undermined the analysis that provided that answer." (Columbia Accident Investigation Board 2003, 160)

"At the January 24, Mission Management Team meeting at which the "no safety-of-flight" conclusion was presented, there was little engineering discussion about the assumptions made, and how the results would differ if other assumptions were used. Engineering solutions presented to management should have included a quantifiable range of uncertainty and risk analysis. " (Columbia Accident Investigation Board 2003, 168)

The Columbia Accident Investigation Board was particularly critical of the frequent use of viewgraphs (e.g. PowerPoint slides) to present overly simplistic representations of issues and analyses.(Columbia Accident Investigation Board 2003, 191).

On Deepwater Horizon, the team conducting the negative pressure tests invented a previously unknown "bladder effect" that would explain the anomalous results that they were seeing. Believing that the tests were accurate would have overturned their previous understanding of the conditions of the well itself, resulting in greater confusion and uncertainty (National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling 2011, 107).

5.5 Blindness can stem from insufficient capacity to perform a probative activity

No safety technique is guaranteed to find all safety problems every time, even when applied competently. There is very limited evidence about the efficacy of safety analysis techniques (Rae, McDermid, and Alexander 2012; Rae, Nicholson, and Alexander 2010), and the evidence that does exist suggests highly variable results depending on who is performing the analysis (Amendola, Contini, and Ziomas 1992).

The prerequisites for a technique to be likely to be effective are that:

- the technique is a match for the subject of the analysis, and used within its range of validity;
- suitable information, including subject matter expertise, is available to populate the technique;
- the technique is applied correctly; and
- the results of the technique are correctly interpreted and applied.

In the absence of these requirements, the output can have the format and language of a correct analysis, but with incorrect content. Worse, preoccupation with getting the format right, particularly where complicated notation is involved, can distract authors and reviewers of safety analysis from the substantive content.

All of the case study accidents involved the use of risk assessment tools requiring specialised knowledge, where lack of that knowledge impeded the analysis.

The Columbia Debris Assessment Team used a software modelling tool called "Crater". The tool was used beyond its ability to validly predict the consequences of the foam strike.

"*Although Crater was designed, and certified, for a very limited set of impact events, the results from Crater simulations can be generated quickly. During STS-107, this led to Crater being used to model an event that was well outside the parameters against which it had been empirically validated."*(Columbia Accident Investigation Board 2003, 144)

As a result of a transfer of responsibilities between Boeing offices, Crater was applied and interpreted by staff who were unfamiliar with its use.

"The engineer had received formal training on Crater from senior Houston-based Boeing engineering staff, but he had only used the program twice before, and had reservations about using it to model the piece of foam debris that struck Columbia (Columbia Accident Investigation Board 2003, 145).

The NTSB recommendations after the 787-8 battery fires cover both initial and recurrent training for engineers who review safety analysis (National Transportation Safety Board 2014, 81). This recommendation was linked explicitly to findings in the report that reviewers had failed to identify and challenge key assumptions made in the safety documents.

The report does not directly question the competence of those who performed the safety analysis, but describes the shortcomings in a way that reflects poorly on their analysis skills.

"Specifically, the analysis did not (1) identify Boeing's assumption that thermal runaway of a cell would not propagate to other cells and (2) provide the engineering rationale needed to justify broad use of this assumption under all operating conditions. Also, the analysis did not sufficiently evaluate and justify the use of the industrial battery failure rate data in predicting the risk of a cell venting occurrence for the 787 battery. Further, even if this information had been included in the EPS safety assessment, the validity of the supporting safety analyses would have been difficult to justify given the limited data available. (National Transportation Safety Board 2014, 71)

The department responsible for the Home Insulation Program had difficulty even finding and keeping project staff.

"DEWHA was subsequently allocated funding to undertake further design and implementation work, but the problem seems to have been not the funding itself, or any lack

of it, but finding suitably qualified and experienced people to assist. [The First Assistant Secretary], for example, said that the level of resources (by which he seemed to mean people) and skill sets available were not commensurate with the tasks allocated."(Hanger 2014, 82)

5.6 Blindness can stem from separation between safety activity and design or operations

For safety analysis to reveal problems with designs or operations, it must first accurately describe them. This requires that those performing the analysis work in close co-operation with those who know most about the system or the relevant work processes.

The Home Insulation Scheme risk assessments did not incorporate expertise with electrical hazards, or listen to concerns raised by electrical trade practitioners.

The failure to engage with the electrical trades reflected an ignorance of one of the major risks attendant upon working in the roof space… it is difficult to understand how a representative of the electrical trades was not invited to subsequent industry consultation meetings. No acceptable explanation was offered for this oversight. (Hanger 2014, 92)

Further, as the scheme evolved, the risk register was not updated to match the changed arrangements. The original scheme envisaged management by a few large contractors, who would also handle training and certification. The final delivery model involved direct payments to homeowners, who could engage small businesses to perform the work.

"There is a real issue whether, given the change to the delivery model that occurred in April 2009, the risk assessment was ever revised to take account of the new and different risks which would arise from a light touch, demand driven and direct engagement model. [The risk management consultant's] evidence was that on 28 April she was told by [a project officer] that the 'business model had moved on significantly since the original workshops were conducted and many of the original risks had become redundant or changed significantly'" (Hanger 2014, 145).

The Columbia Accident Investigation Board adopted a term coined by the earlier Roger Commission into the Challenger Accident, the "Silent Safety Program". This referred to the disconnect between safety activities and operational decision making.

"The Debris Assessment Team, working in an essentially decentralized format, was well-led and had the right expertise to work the problem, but their charter was "fuzzy," and the team had little direct connection to the Mission Management Team. This lack of connection to the Mission Management Team and the Mission Evaluation Room is the single most compelling reason why communications were so poor during the debris assessment" (Columbia Accident Investigation Board 2003, 180).

GS Yuasa was very experienced with lithium batteries, but had limited knowledge about aircraft design or operations:

"GS Yuasa did not adequately account for a number of factors that were relevant to propagation risk. For example, the test was not conducted at the battery's maximum operating temperature of 158ºF, and the test setup did not fully represent the battery installation on the 787 airplane…Further, the test was performed using a development unit that did not incorporate the final battery design certified as part the 787 type design." (National Transportation Safety Board 2014, 69)

5.7 Blindness can arise when it is hard to respond to any problems that are found

Trade-offs between efficiency and thoroughness (Hollnagel 2009) are a well understood source of variation in work performance, including the execution of safety tasks. The less time that is available to complete a task, the more emphasis shifts towards completing the task rather than addressing every detail.

From the case studies, it is clear that thoroughness is also adversely affected if there is insufficient resource and time to fix any newly identified issues. The task may still be performed diligently, but effort is redirected towards the support of existing beliefs rather than discovery of new information. This is particularly the case if finding problems contradicts previous assurances about the integrity of the design, or causes a loss of face or reputation.

The safety analysis is often performed once decisions about the system or operations have already been made, and when there would be considerable financial or reputational cost invoked if new safety issues were identified. It is common practice, for example, to create "retrospective" safety cases for equipment currently in service (Aas, Andersen, and Skramstad 2009; Eriksson 2004; Hill 2007). As the review into the Nimrod XV230 crash highlighted, these are often exercises in confirming institutional assumptions about safety. Shortly prior to the crash of XV230 in Afghanistan, the Nimrod aircraft was the subject of a "legacy" safety case. This was viewed in advance as an opportunity to create structured evidence to support "the high level of corporate confidence in the safety of the Nimrod aircraft", and with hindsight as a missed opportunity to assess the risk of an extensively modified and aging airframe. "The problem was that those involved in producing the NSC [Nimrod Safety Case] embarked on the process believing the Nimrod type was safe" (Haddon-Cave 2009, 189). The focus was on demonstrating what was already "known", with less attention paid to the need to discover and communicate risk.

For the Home Insulation Program, fixed deadlines provided an incentive not to raise issues that would require further changes to the scheme.

"What is abundantly clear from the evidence is that from the outset the 1 July 2009 commencement date was thought to be non-negotiable by those public servants and consultants working on the program." (Hanger 2014, 86)

There were also concerns about allowing internal industry disputes to disrupt the smooth delivery of the program.

"[A representative for an installation installer] believes the comments he made arose in the context of safety. He said that the Chairman, [the assistant secretary], discouraged discussion at that time. He said that [the assistant secretary] said something to the effect of 'okay, we will note that and move on'.

The Minutes do not reflect his recollection of how the meeting proceeded. He said he had been told not to 'rock the boat' at these meetings. This was an impression he had obtained from [the assistant secretary], namely that he did not want internal industry disputes raised. It was not a comfortable situation to be in, [a representative for an installation installer] said, because he was 'rocking the boat'.

[The assistant secretary], for his part, was no doubt trying to control what was a fragmented and divisive industry." (Hanger 2014, 94).

These problems were exacerbated by the fact that the Home Insulation Program used an integrated risk management approach that focused primarily on political and program risks.

Where safety was an issue it was typically viewed through the lenses of program delays and political fallout. Concerns such as fraud and training were raised by some participants as safety issues, but were understood by others to have no safety implications.

"The real difficulty is that [the Parliamentary Secretary] seems to have formed a view that participants in the HIP might act fraudulently or dishonestly, but not had an understanding that this might readily extend to a serious failure to adhere to their occupational health and safety obligations." (Hanger 2014, 124)

Delaying a space shuttle mission was seen as an embarrassing failure.

"Further, when asked by investigators why they were not more vocal about their concerns, Debris Assessment Team members opined that by raising contrary points of view about Shuttle mission safety, they would be singled out for possible ridicule by their peers and managers." (Columbia Accident Investigation Board 2003, 169)

The accident investigation board saw this as part of a pattern where deadlines placed pressure not just on the time available for safety activities, but on the acceptability of bad news arising from those activities.

"Ultimately, external expectations and pressures impact even data collection, trend analysis, information development, and the reporting and disposition of anomalies. These realities contradict NASA's optimistic belief that pre-flight reviews provide true safeguards against unacceptable hazards." (Columbia Accident Investigation Board 2003, 190)

Rather than being seen as evidence of diligence, newly raised safety issues were perceived as signs of disorganisation and ineptitude.

"*It is here that the decision to fly before resolving the foam problem at the STS-113 Flight Readiness Review influences decisions made during STS-107. Having at hand a previously accepted rationale – reached just one mission ago – that foam strikes are not a safety-of-flight issue provides a strong incentive for Mission managers and working engineers to use that same judgment for STS-107. If managers and engineers were to argue that foam strikes are a safety-of-flight issue, they would contradict an established consensus that was a product of the Shuttle Program's most rigorous review – a review in which many of them were active participants."* (Columbia Accident Investigation Board 2003, 150)

The rigor of the Federal Aviation Administration's certification process created inertia that worked against the interests of safety. Achieving sign-off of a safety deliverable was an arduous and expensive process that would be "undone" by identifying a new problem. Even the plan for producing the safety analysis was a pre-approved deliverable. Challenging the analysis would have required claiming that the analysis process – already agreed by all parties – was inadequate.

Thus, the FAA could not effectively use traceability principles to evaluate the completeness of Boeing's proposed methods of compliance, particularly for special condition 2, which addressed battery thermal runaway. As a result, the FAA approved Boeing's proposed EPS certification plan, including qualification tests, for the 787 main and APU battery without the details necessary to demonstrate compliance with the individual special conditions. (National Transportation Safety Board 2014, 74)

6 Conclusions

This paper discusses four major accidents in depth. These are by no means the only instances of probative blindness identified by the authors. The history of accidents is replete with discussion of flawed or ignored safety analysis, and missed opportunities to identify and mitigate hazards.

It is tempting to assume that pre-accident safety analysis is in some way exceptional – that these "failed" safety activities are unlike other "normal" safety activities. Unfortunately, the circumstances that lead to probative blindness are not exceptional. They are conditions that prevail in many, if not the majority of organisations working with safety critical systems. It is common to find any of the conditions from Figure 2:

- strong prior beliefs about safety arising from operational experience or from previous safety analysis
- transfer of concern from safety to compliance
- "successful" safety framed as not finding problems
- a desire for safety activities to provide certainty and closure
- techniques that require specialised notation, skills or domain knowledge, with difficulty finding and keeping staff competent in these techniques
- separation between safety activities and design or operations
- safety activities conducted when there is no time or budget to respond to negative findings, or where there is loss of reputation associated with admitting safety problems

If these are the circumstances that give rise to probative blindness – and our study strongly suggests that they are – then there is good reason to look upon any safety activity with a skeptical eye. Do these activities constitute a genuine search for knowledge about safety risk, or are they responses to regulatory, organisational and individual factors that demand certainty and assurance?

Testing our conclusions will require extending the investigation to cover safety activities in the absence of accidents. Non-accident operation provides a larger and more neutral environment for examining safety activities, and allows direct manipulation of techniques and the conditions under which they are applied. It is also this environment in which explanatory theories of probative blindness may ultimately be applied to avoid accidents. Probative blindness in non-accident operation will be revealed not by the occurrence of disaster, but by the absence of changed belief. How often do safety activities actually result in significant design or operational changes? Where changed belief does occur, is it a consequence of the techniques that are used, the people who apply them, or the context in which they are applied? Is probative blindness a symptom of certain pathologies, and can they be treated?

7 References

Aas, Andreas Lumbe, Heidi Stenberg Andersen, and Torbjorn Skramstad. 2009. "A Retrospective Safety Case for an Advanced Driller's Cabin Using the Goal Structuring Notation (GSN)." In *International Petroleum Technology Conference*. Doha. doi:10.2523/13755-MS.

Amendola, A., S. Contini, and I. Ziomas. 1992. "Uncertainties in Chemical Risk Assessment: Results of a European Benchmark Exercise." *Journal of Hazardous Materials* 29 (3): 347–63. doi:10.1016/0304-3894(92)85041-X.

Apostolakis, George E. 2004. “How Useful Is Quantitative Risk Assessment?” *Risk Analysis* 24 (3): 515–20. doi:10.1111/j.0272-4332.2004.00455.x.

Barber, Edward Hamilton. 1971. “Report of Royal Commission into the Failure of West Gate Bridge.” 2–7037/71. Melbourne: Government Printer. http://www.westgatebridge.org/sites/default/files/downloads/report-of-royal-commission.pdf.

Barton, N., and A.J. Rae. 2012. “Unplugged Perils, Lost Hazards and Failed Mitigations.” In *7th IET International Conference on System Safety, Incorporating the Cyber Security Conference 2012*, 1–6. doi:10.1049/cp.2012.1496.

Besnard, Denis, and Erik Hollnagel. 2012. “I Want to Believe: Some Myths about the Management of Industrial Safety.” *Cognition, Technology & Work*, 1–11. doi:10.1007/s10111-012-0237-4.

Board, Major Incidents Investigation. 2006. “Buncefield Investigation.” Website facility. February 21. http://www.buncefieldinvestigation.gov.uk/reports/index.htm#final.

Bogen, James, and James Woodward. 1988. “Saving the Phenomena.” *The Philosophical Review* 97 (3). http://www.pitt.edu/~rtjbog/bogen/saving.pdf.

BP. 2010. “Deepwater Horizon Accident Investigation Report.” http://www.bp.com/content/dam/bp/pdf/sustainability/issue-reports/Deepwater_Horizon_Accident_Investigation_Report.pdf.

Columbia Accident Investigation Board. 2003. “Report.” National Aeronautics and Space Administration. http://www.nasa.gov/columbia/home/CAIB_Vol1.html.

Corbin, Juliet M., and Anselm Strauss. 1990. “Grounded Theory Research: Procedures, Canons, and Evaluative Criteria.” *Qualitative Sociology* 13 (1): 3–21. doi:10.1007/BF00988593.

Deepwater Horizon Study Group. 2011. “Final Report on the Investigation of the Macondo Well Blowout.” University of California. http://ccrm.berkeley.edu/pdfs_papers/bea_pdfs/dhsgfinalreport-march2011-tag.pdf.

Ditto, Peter H., and David F. Lopez. 1992. “Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions.” *Journal of Personality and Social Psychology* 63 (4): 568–84. doi:10.1037/0022-3514.63.4.568.

Dombrowski, P.M. 1991. “The Lessons of the Challenger Investigations.” *Professional Communication, IEEE Transactions on* 34 (4): 211–16. doi:10.1109/47.108666.

EQE International. 2000. “Risk Assessment of the Deepwater Horizon Blowout Preventer (BOP) Control System.” EQE Project Number 253150. Cameron Controls Corp. http://mdl2179trialdocs.com/releases/release201303211200016/TREX-04275.pdf.

Eriksson, Lars-henrik. 2004. “Using Formal Methods in a Retrospective Safety Case.” In *Computer Safety, Reliability, and Security – 23rd International Conference, SAFECOMP 2004, Springer Lecture Notes in Computer Science 3219, Springer-Verlag*.

Fischhoff, B. 2003. “Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment under Uncertainty.” *Quality and Safety in Health Care* 12 (4): 304–11. doi:10.1136/qhc.12.4.304.

Haddon-Cave, Charles. 2009. “The Nimrod Review: An Independent Review into the Broader Issues Surrounding the Loss of the RAF Nimrod MR2 Aircraft XV230 in Afghanistan in 2006.”

Hanger, Ian. 2014. “Report of the Royal Commission into Home Insulation Program.” Royal Commission into the Home Insulation Program.

Hill, J. 2007. “A Software Safety Risk Taxonomy for Use in Retrospective Safety Cases.” In *31st IEEE Software Engineering Workshop, 2007. SEW 2007*, 179–86. doi:10.1109/SEW.2007.50.

Hollnagel, Erik. 2009. *The ETTO Principle: Efficiency-Thoroughness Trade-Off*. Burlington, VT: Ashgate.

International Nuclear Safety Advisory Group. 1991. “Safety Culture.” INSAG-4. Safety Series. Vienna: International Atomic Energy Agency.

Japan Transport Safety Board. 2014. “Aircraft Serious Incident Investigation Report: All Nippon Airways Co., Ltd., JA804A.” AI2014-4. JTSB.

Kewell, Beth. 2007. “Linking Risk and Reputation: A Research Agenda and Methodological Analysis.” *Risk Management* 9 (4): 238–54.

Keyser, V. De, and D. D. Woods. 1990. “Fixation Errors: Failures to Revise Situation Assessment in Dynamic and Risky Systems.” In *Systems Reliability Assessment*, edited by A. G. Colombo and A. Saiz de Bustamante, 231–51. ISPRA Courses 6. Springer Netherlands. http://link.springer.com/chapter/10.1007/978-94-009-0649-5_11.

La Porte, Todd R. 1996. “High Reliability Organizations: Unlikely, Demanding and At Risk.” *Journal of Contingencies & Crisis Management* 4 (2): 60.

Menon, Cather ine, Richard Hawkins, and John McDermid. 2009. “Defence Standard 00-56 Issue 4: Towards Evidence-Based Safety Standards.” In *Safety-Critical Systems: Problems, Process and Practice*, edited by Chris Dale and Tom Anderson, 223–43. London: Springer London. http://www.springerlink.com/content/p233v62x27m651q1/.

National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling. 2011. “Deep Water: The Gulf Oil Disaster and the Future of Offshore Drilling: Report to the President.” Washington, D.C.: US GPO.

National Transportation Safety Board. 2014. “Auxiliary Power Unit Battery Fire Japan Airlines Boeing 787-8, JA829J.” Incident Report NTSB/AIR-14/01. NTSB.

Perrow, Charles. 1999. *Normal Accidents: Living with High-Risk Technologies*. Princeton University Press.

Rae, A. J., Rob Alexander, and John McDermid. 2014. “Fixing the Cracks in the Crystal Ball: A Maturity Model for Quantitative Risk Assessment.” *Reliability Engineering & System Safety* 125 (May): 67–81. doi:10.1016/j.ress.2013.09.008.

Rae, A. J., J. A. McDermid, R. D. Alexander, and M. Nicholson. 2014. “Probative Blindness: How Safety Activity Can Fail to Update Beliefs about Safety.” In *System Safety and Cyber Security*. Manchester: IET.

Rae, A. J., J.A McDermid, and R. D Alexander. 2012. “The Science and Superstition of Quantitative Risk Assessment.” In *Annual European Safety and Reliability Conference*. Helsinki.

Rae, A. J., M. Nicholson, and R.D. Alexander. 2010. “The State of Practice in System Safety Research Evaluation.” In *IET System Safety Conference*. Manchester.

Turner, Barry A. 1976. “The Organizational and Interorganizational Development of Disasters.” *Administrative Science Quarterly* 21 (3): 378–97. doi:10.2307/2391850.

U.S. Chemical Safety and Hazard Investigation Board. 2014. “Explosion and Fire at the Macondo Well.” 2010–10–I–OS. Government Printer. http://www.csb.gov/assets/1/7/Overview_-_Final.pdf.

Vaughan, Diane. 1997. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. 1 edition. Chicago: University Of Chicago Press.

Vaughan, Diane. 2004. ‘Theorizing Disaster: Analogy, Historical Ethnography, and the Challenger Accident’. Ethnography 5 (3): 315–47. doi:0.1177/1466138104045659.

von Krogh, Georg, Cristina Rossi-Lamastra, and Stefan Haefliger. 2012. “Phenomenon-Based Research in Management and Organisation Science: When Is It Rigorous and Does It Matter?” *Long Range Planning* 45 (4): 277–98. doi:10.1016/j.lrp.2012.05.001.

Wastell, David G. 1996. “The Fetish of Technique: Methodology as a Social Defence.” *Information Systems Journal* 6 (1): 25–40.

Weick, Karl E. 1993. “The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster.” *Administrative Science Quarterly* 38 (4): 628–52. doi:10.2307/2393339.

Yin, Robert K. 2003. *Case Study Research: Design and Methods: (Applied Social Research Methods, Volume 5): 005*. Third Edition. Sage Publications, Inc.